Effective Reconstruction of Backprojection images through Attention Mechanism

Venkata Chowdary[1] , Venkata Sai Hithesh Reddy[1] , Thejeshwar Reddy[1] , Sunil Kumar[1] , Rajasekaran.M[1*]

[1] Department of Computer Science and Engineering,

Madanapalle Institute of Technology and Science, Madanapalle – 517325, INDIA.

Abstract. Compared to time-domain photoacoustic imaging, frequency-domain photoacoustic (FDPA) imaging has much more potential in a clinical setting because of its smaller size and lower cost. elements. However, because of its poorer signal-to-noise ratio, the FDPA system requires sophisticated image reconstruction techniques. In FDPA imaging, most image reconstruction techniques rely on analytical or model-based schemes [1]. This work developed an image reconstruction technique based on deep learning that can directly reconstruct backprojection images and enhance their quality. This architecture was inspired by U Net, which uses attention gates at the skip connections and a sequence of encoders and decoders after that. A comparison is made between the outcomes and direct translational networks built on vanilla U Net. By using our proposed model, we observed an improvement of about 10% on both PSNR and SSIM metrics.

Keywords: photoacoustic imaging, frequency domain, Deep Learning, U-Net, Attention

1   Introduction

Over the past decade, photoacoustic tomography (PAT) has become a highly effective imaging modality with a wide range of uses in brain, breast, and vasculature imaging. Using this novel technique, broadband acoustic waves are produced when tissue expands thermoelastically in response to near-infrared light absorption. Reconstructing the initial pressure distribution of the acoustic source is a necessary step in the PAT image reconstruction process. Because analytical methods are computationally efficient, even at the expense of image quality, they have historically been used for image reconstruction. Nonetheless, recent developments have led to incorporating deep learning-based post-processing techniques for PAT-limited data scenarios and artifact removal. Notably, models such as U-Net[4] have become increasingly well-liked due to their efficacy in various image segmentation tasks. In our study, to improve the quality of image reconstruction in photoacoustic tomography, we employ an attention U-Net model-based approach.

2   Dataset and Pre-Processing

We used the numerical breast phantoms from the OA-Breast Database [2] in this work. Three unique body phantoms make up this database; each voxel represents a different form of tissue, such as the background, vasculature, skin, fat, and fibro-glandular tissues[10]. The backdrop is set to zero and each tissue type is assigned a specific value in kilopascals (kPa). The simulations started with the assumption of a constant optical fluence and only considered the first induced pressure. Later, non-uniform optical fluence was added for added realism [5][6]. We created our dataset by slicing the 3D breast phantom to produce 2D images to make computation more feasible, all slices were resized to 128 × 128 pixels. To mimic photoacoustic tomography imaging[17], we added an additional noise of SNR 40 to our images. The calculations only considered the first induced pressure and were predicated on the idea of a continuous optical fluence. For extra realism, non-uniform optical fluence was later incorporated; further details are provided in the following sections.

3   Evaluation metrics

To assess the performance of the convolution network in reconstructing ground truths from backprojection, we employ two distinct evaluation metrics in this study.

1) Peak Signal to Noise Ratio (PSNR): PSNR depicts the quality of the reconstructed by comparing the mean squared error (MSE) between the ground truths (Xa) and the reconstructed image (Xa hat) on a logarithmic scale. It is calculated using the formula:

$$PSNR = 20\,log_{10}(MAXI\,\sqrt{MSE})$$

where MSE is the mean squared error computed as:

$$MSE = \frac{1}{n^2}\sum_{i,j=1}^{n}\left(X_{aij} - \widehat{X}_{aij}\right)$$

The maximum intensity of the reconstructed age is denoted by MAXI in this instance. As the quality of the reconstructed image improves, PSNR rises.

2) Structural Similarity Index Measure (SSIM): SSIM assesses the features similarity and structural similarity between the reconstructed image and the ground truth. It yields a value between -1 to 1. A higher SSIM value signifies better quality of the reconstruction.

4   Training and Implementation Details

To improve the quality of the reconstructed images, we first generated initial reconstructions from the limited acoustic boundary data using an analytical backprojection technique. We then segmented tissue types and improved the reconstructed images using an image reconstruction method based on the attention U-Net model. We utilized an attention-based neural network architecture (which includes depth-wise convolutions in its encoder-decoder structure) [3]. This neural network segmented the image into classes that corresponded to the different tissue types previously mentioned.

Following the first back-projection reconstruction, 1731 images in all were obtained. We used data augmentation methods to add even more to our dataset. The dataset was then segregated into two groups: a validation set with 860 (20%) images and a training set with 3318 (80%) images. Using Adam as the optimizer and the peak signal-to-noise ratio as the loss function

4.1   Proposed Model

To reconstruct the analytical backprojection images to their original form, many experiments are conducted with vanilla U-Nets [1], but none of them focused on integrating the Attention mechanism to their models. Attention mechanism was previously shown to enhance the image quality to a greater degree by forcing the model to look at a particular area of the image in our case, we tuned our attention model to focus more on the fibro-glandular tissues which happens to be the most important and often neglected part in backprojection process. Due to the integration of attention mechanism, the model converges much faster and has better generalization capabilities [10].
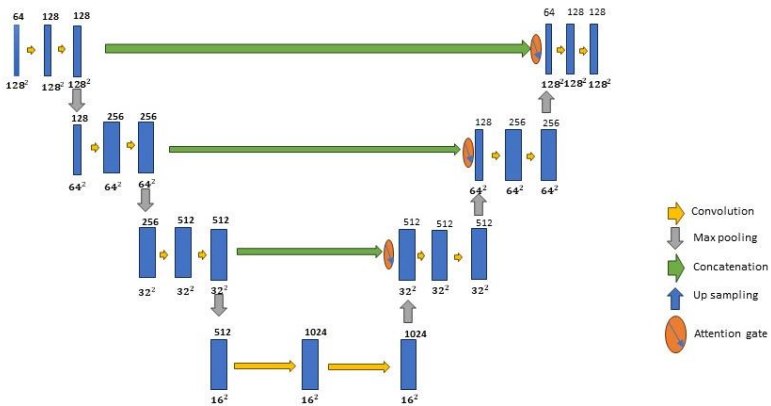
We integrated this attention mechanism as attention gates as shown in Fig.2 at the skin connections by concatenating both together and then adding it to the decoder path in various stages. Our model is an amalgamation of both vanilla U-Nets and Attention. At the encoder phase the backprojection images of dimensions $128 \times 128 \times 1$ will be convoluted with filters of size 3×3, has 64 channels has a stride value as '2'. The output of the first convolution part is given to the second part of the encoder which has the same hyperparameters as the previous layer and additionally max pooling is applied at the start of the second part of the encoder. The same process runs for one mode encoder step. Through this process the model will learn the essential features of the images through feature extraction. At the bottleneck part of the model the images will be of dimension $16 \times 16 \times 512$ with a reduction of $32 \times 32 \times 512$ and $64 \times 64 \times 128$ respectively as shown in the Fig.1a. The final layer of encoding part is then be submitted as input to the first layer of decoder part with a dimension of $16 \times 16 \times 1024$. The decoder part of the network consists of deconvolution layers which will learn to decode the features and reconstruct the image. The dimensions are repeated so that the output part of the decoder will give the same dimensions as the input layer. Throughout this whole process the network will lack the spatial features of the input images which are essential

for image reconstruction. To mitigate this issue, we add skip connections directly from the encoder part to decoder part. As we are targeting the improvement of the spatial features, we are adding our attention gate to the skip connections.

The attention gate mechanism uses data from a neural network's shallow and deep layers to improve feature representation [3]. It enhances aligned features and suppresses unaligned ones by adding elements at a time from input vectors x and g. Following its passage through convolution layers and ReLU activation, the resultant vector is subjected to sigmoid scaling, which yields attention coefficients. These coefficients serve as a guide for the trilinear interpolation up sampling process by indicating the significance of the features. Ultimately, the coefficients are applied to the original x vector, highlighting important features and allowing the network to concentrate on relevant data, improving performance by extracting features selectively.

Together with all the hyperparameters, Our Attention U-Net model was trained for 50 epochs. Notably, by using segmentation assignment to recover each tissue type found in the ground truth image, a combined reconstruction was produced as the result, which enhanced the overall precision and quality of the reconstructed images.

Fig. 1. Attention U-Net Arcitecture

5　　Results

Two sets of results predicted with our model for the numerical breast phantom figures are shown in Figure 2. Here, the image reconstructed from the vanilla U-Net shows the entire reconstruction process, while the image designated as the "Reconstructed image through Attention U-Net" image represents the proposed model. This limited data analytical backprojection yields the backprojection image, which is then fed into the U-Net architectures.
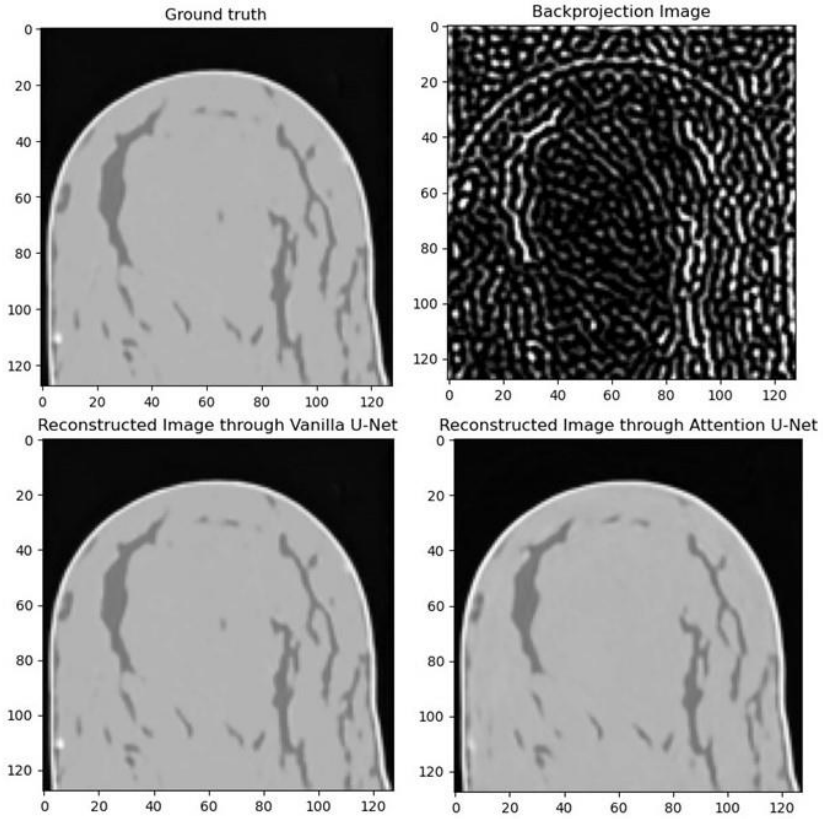
Fig. 2. Top, Ground Truth and Image Simulated by using Backprojection. Bottom, Reconstructed image through Vanilla U-Net and Reconstructed Image through Attention U-Net.

Architecture	SSIM	PSNR
Vanilla U-Net	0.7576	28.6275
Attention U-Net	0.8489	30.3554

Table 1. PSNR and SSIM values of Vanilla U-Net and Attention U-Net

We employed 2 different metrics to assess our model's performance on the test set of images. We obtained a PSNR- Peak signal to noise ratio across various classes (tissue categories) and SSIM, demonstrating the efficacy of our method in precisely segmenting different tissue types in the reconstructed images. This measure emphasizes how well the model can mimic tissue structures and shows how it can improve the quality of image reconstruction in photoacoustic tomography applications. Through our

repetitive experiments, we got an SSIM and PSNR of 0.8489 and 30.3554 which out-performs the traditional U-Net which gives an SSIM and PSNR of 0.7576 and 28.6274 as represented in Table.1. So, through these results we can conclude that Attention U-Net works better than Vanilla U-Nets in Image Reconstruction tasks.

6   Conclusion and Future Scope

With this study, it is evident that by including attention mechanism to an autoencoder model like U-Net, we were able to increase the performance of the existing model by almost 10%. With these results in mind, we would like to continue experimenting with other popular architectures like Vision Transformers (ViT) which is one among the popular encoder-decoder based architectures in computer vision and has multi-head attention enabled. In addition to that we will also continue experimenting with popular segmentation algorithms like UNETR (UNet Transformer) to reconstruct analytical models to a greater clarity which hopefully will yield better PSNR and SSIM values.

References

1. Hauptmann, A., & Cox, B. (2020). Deep learning in photoacoustic tomography: current approaches and future directions. *Journal of Biomedical Optics*, *25*(11), 112903-112903.
2. Lou, Y., Zhou, W., Matthews, T. P., Appleton, C. M., & Anastasio, M. A. (2017). Generation of anatomically realistic numerical phantoms for photoacoustic and ultrasonic breast imaging. Journal of biomedical optics, 22(4), 041015-041015.
3. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., ... & Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
4. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.
5. Aldridge, S., & Downs, A. J. (2001). Hydrides of the main-group metals: New variations on an old theme. *Chemical reviews*, *101*(11), 3305-3366.
6. Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. *Journal of Avian Biology*, *49*(5), jav-01447.
7. Incze, A., Jancsó, H. B., Szilágyi, Z., Farkas, A., & Sulyok, C. (2018, September). Bird sound recognition using a convolutional neural network. In *2018 IEEE 16th international symposium on intelligent systems and informatics (SISY)* (pp. 000295-000300). IEEE.
8. Mai, L., Bao, L. J., Shi, L., Wong, C. S., & Zeng, E. Y. (2018). A review of methods for measuring microplastics in aquatic environments. *Environmental Science and Pollution Research*, *25*, 11319-11332.
9. Poluha, R. L., Canales, G. D. L. T., Costa, Y. M., Grossmann, E., Bonjardim, L. R., & Conti, P. C. R. (2019). Temporomandibular joint disc displacement with reduction: a review of mechanisms and clinical presentation. *Journal of applied oral science*, *27*, e20180433.

10. Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K. H., & Frommolt, K. H. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, *31*(12), 1524-1534.

11. Martinsson, J., & Lundqvist, L. J. (2010). Ecological citizenship: coming out 'clean'without turning 'green'?. *Environmental politics*, *19*(4), 518-537.

12. Brandes, T. S. (2008). Automated sound recording and analysis techniques for bird surveys and conservation. *Bird Conservation International*, *18*(S1), S163-S173.

13. Wu, T., Qin, Z., Wang, Y., Wu, Y., Chen, W., Zhang, S., ... & Han, L. (2021). The main progress of perovskite solar cells in 2020–2021. *Nano-Micro Letters*, *13*, 1-18.

14. Kocsis, P., Súkeník, P., Brasó, G., Nießner, M., Leal-Taixé, L., & Elezi, I. (2022). The unreasonable effectiveness of fully-connected layers for low-data regimes. *Advances in Neural Information Processing Systems*, *35*, 1896-1908.

15. Deán-Ben, X. L., Ding, L., & Razansky, D. (2017). Dynamic particle enhancement in limited-view optoacoustic tomography. *Optics Letters*, *42*(4), 827-830.

16. Schmalz, J. A., Schmalz, G., Gureyev, T. E., & Pavlov, K. M. (2010). On the derivation of the Green's function for the Helmholtz equation using generalized functions. *American Journal of Physics*, *78*(2), 181.

17. A. Vinaya Babu, and S. Viswanadha Raju. "Clustering of Concept-Drift Categorical Data Implementation in JAVA." In International Conference on Computing and Communication Systems, pp. 639-654. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

18. Kellnberger, S., Rosenthal, A., Myklatun, A., Westmeyer, G. G., Sergiadis, G., & Ntziachristos, V. (2016). Magnetoacoustic sensing of magnetic nanoparticles. *Physical Review Letters*, *116*(10), 108103.

19. Treeby, B. E., & Cox, B. T. (2010). Modeling power law absorption and dispersion for acoustic propagation using the fractional Laplacian. *The Journal of the Acoustical Society of America*, *127*(5), 2741-2748.